

國立臺灣師範大學管理學院全球經營與策略研究所

碩士論文

Graduate Institute of Global Business and Strategy

College of Management

National Taiwan Normal University

Master's Thesis

採用隱含狄利克雷分佈

分析使用者評論：以 Steam 網站為例

Application of Latent Dirichlet Allocation on Analyzing

User Reviews: Case Study on Steam Website

王哲宏

Jhe-Hong Wang

指導教授：施人英 博士

Advisor: Jen-Ying Shih, Ph.D.

中華民國112年2月

February 2023

謝誌

求學的時光總是過得特別快，研究所階段也接近尾聲，即將邁入人生的下個階段。在寫作論文的過程中，首先特別感謝指導教授施人英老師，在論文選題上給予很大的自由度，同時也對於論文內容給予很多技術相關上的建議，讓我能順利完成論文及口試。同時也感謝口試委員陳文華教授與何宗武教授針對論文內容也給予許多寶貴意見。而何教授更是讓我開始接觸程式語言的契機，讓我在研究所學到了過去較少接觸到的領域。

其次，在就讀研究所期間，也特別感謝我的朋友們，平時一起聚會聊天都可以讓我紓解緊繃的心情。踏入研究所之後，遇到許多值得感謝的人，即使遇到許多挫折及迷惘，也還是努力走完求學路，我打從心底非常感謝所有曾經幫助過我的人。

最後，也非常感謝我的家人，從小到大都支持我做想做的事，讓過毋須煩惱課業以外的事，寫論文期間也不斷給予我安慰與鼓勵，真的非常感謝默默支持我的家人們。謹此將論文獻給我的家人以及過去所有幫助與鼓勵我的人。

王哲宏 謹誌 2023.2

摘要

過去的研究大多著重於檢驗電子遊戲專家評論，僅有少數研究在探索遊戲玩家提供的使用者生成內容(UGC)。使用者生成內容包含玩家感受的豐富資訊，有助於遊戲開發商了解玩家的真實想法。基於使用者生成內容，本研究著重於分析玩過角色扮演遊戲(Role-Playing Game，簡稱RPG)玩家的使用者生成內容，以探索他們的想法和態度。

本研究提出一個適用於遊戲評論文本主題的分析方法，主要是依據隱含狄利克雷分布(Latent Dirichlet Allocation，簡稱 LDA)演算法，它可以將文本集中成一篇文章，而每篇文本的主題則按照概率分布的形式。從 Steam 遊戲平台所擷取的玩家評論內容中或許含有關於角色、遊戲畫面、故事等相關的評論主題，這些主題可以視為玩家的特徵，因此本研究運用 LDA 主題模型，從玩家的遊戲評論內容分析出玩家特徵以及遊戲的關鍵屬性。

綜合研究結果可以確認，本研究選擇的三款目標遊戲雖都同屬於角色扮演遊戲類型，但遊戲內容及玩法不盡相同(如故事背景、遊戲視角等)。本研究將評論資料基於 LDA 主題模型分析出大部分玩家著重於對於角色扮演遊戲的規則及目標，但由於三款遊戲還是有不同內容，因此從 LDA 主題模型分析出也會有相異的主題屬性。

本研究提供遊戲開發商一個實務方式來有效率確認玩家關注的主題，對於遊戲的後續改善提出建議，以及開發新遊戲提出方向。

關鍵字：使用者評論、主題模型、隱含狄利克雷分布、Steam 遊戲平台

Abstract

In the past, most of the studies focused on game reviews from experts, with only a few exploring the reviews provided by players. User reviews contain a wealth of information about players' feelings and helps developers understand players' thoughts. Based on user-generated content, this study focuses on analyzing the user-generated content of role-playing games players to explore their thoughts and attitudes.

This study proposes an analysis method suitable for game reviews, mainly based on the Latent Dirichlet Allocation(LDA), which can collect all the reviews into one document, and the topic of each is in the form of probability distribution given. The player review contents extracted from Steam website may contain some topics such as characters, graphics, stories, etc. These topics can be considered as the characteristics of the player. Therefore, this study uses the LDA topic model to analyze the player's reviews and analyzes players' characteristics and key attributes of the games.

Based on the results, it can be confirmed that although the three different games selected are role-playing games, however the game content and gameplay are different (such as story background, game perspective, etc.). This study analyzes the review based on the LDA topic model and finds that most players focus on the rules and objectives of role-playing games, but since these three games still have different contents, the LDA topic model will also have different topic attributes.

This study provides an empirical method for developers to identify the topics that players focus on and gives suggestions for the following improvement.

Keywords: User Review, Topic Model, Latent Dirichlet Allocation, Steam

目錄

第一章、緒論.....	1
1.1 研究背景.....	1
1.2 研究動機.....	3
1.3 研究架構.....	3
第二章 文獻探討.....	5
2.1 使用者生成內容與評論的應用.....	5
2.2 主題模型與隱含狄利克雷分布.....	6
2.3 遊戲屬性分類.....	8
第三章 研究方法.....	10
3.1 資料蒐集.....	10
3.2 文字預處理.....	11
3.3 主題模型應用.....	13
第四章 研究結果.....	16
4.1 研究資料.....	16
4.2 結果評估.....	16
4.2.1 電馭叛客 2077.....	16
4.2.2 巫師 3：狂獵.....	17
4.2.3 隻狼：暗影雙死.....	19

4.3 綜合討論.....	20
第五章 結論與建議.....	21
5.1 結論.....	21
5.2 研究限制.....	21
5.3 未來研究方向.....	22
參考文獻.....	24
附錄.....	28



圖目錄

圖 1.1 研究架構圖.....	4
圖 2.1 主題模型示意圖.....	6
圖 3.1 Steam 評論之使用者介面.....	11
圖 3.2 遊戲評論資料示意圖.....	12



表目錄

表 2.1 主題模型研究相關整理.....,,,	7
表 2.2 遊戲屬性說明.....	9
表 3.1 三款 RPG 遊戲所對應之評論數量.....	10
表 3.2 停止詞列表.....	13
表 3.3 各遊戲最適主題數量.....	14
表 4.1 LDA 設定參數.....,,,	16
表 4.2 電馭叛客 2077 之分布佔比前五個主題列表.....	17
表 4.3 巫師 3：狂獵之分布佔比前五個主題列表.....	18
表 4.4 隻狼：暗影雙死之分布佔比前五個主題列表.....	19
表 4.5 三款遊戲之屬性標籤.....	20

第一章 緒論

1.1 研究背景

隨著網路及電子硬體設備的快速發展，大眾對於電子娛樂有愈來愈多種選擇，其中包括數位影音串流、社群媒體、電子遊戲等，近年來需求逐年上升。近期，因去年新冠肺炎疫情(COVID-19 Pandemic)的居家隔離，更是帶動電子遊戲的需求大幅成長(Facebook Gaming, 2021)。

以電子遊戲產業規模，遊戲開發團隊要成功開發一款暢銷且口碑良好的遊戲是非常具有挑戰性。根據先前的研究，相較於一般的消費者，遊戲玩家相比於一般消費者是較難被滿足(Lin, Bezemer, Zou & Hassan, 2019)，這使得遊戲玩家對於一款遊戲的態度及遊戲的品質會成為影響該遊戲銷量的重要因素。為了提高遊戲品質，遊戲開發團隊是否能了解遊戲玩家的體驗是非常重要的，而使用者評論正是代表遊戲玩家感受的直接體現。

對於使用者評論的使用相當廣泛，像是 Apple 的 App Store 和 Google Play 等手機應用程式平台。這些 App 評論除了提供了豐富的數據來源，也可以用來了解使用者的心得以及 App 時所發生的問題。過去有關手機 App 評論的研究(吳尚緯, 2013)表明了研究使用者評論的價值。

在電子遊戲產業，遊戲評論已成為遊戲玩家與遊戲開發團隊間最直接的互動方式。透過大量的玩家，持續累積每款遊戲的評論數量，從而產生了龐大的遊戲社群。遊戲開發團隊就能從社群內找出更多玩家的想法，而遊戲評論是更玩家想法的直接反映，因此對評論的更好的理解可以協助遊戲開發團隊開發具有更高品質的遊戲(Lin et al., 2019)。

龐大的電子遊戲產業規模包括大量的遊戲玩家，他們提供了大量關於電子遊戲的使用者生成內容。較為常見的內容形式像網路論壇如 Reddit、PTT 或遊戲平台如 Steam 等進行評論。使用者評論包含了遊戲玩家對該款遊戲最真誠的看法以及諸多玩家常提到的功能。通過對評論的分析，遊戲

開發團隊可以獲得遊戲玩家豐富、大量及最直接的使用者生成內容，有利於遊戲改進和未來推出的下一款遊戲。

而目前著名的遊戲社群網路服務平台 Steam，是由美國電子遊戲開發商 Valve 於 2003 年推出的遊戲數位發行平台，也是目前全世界的遊戲平台中，遊戲數量最多、功能最為完善、使用者體驗最為良好的平台。該平台同時提供數位版權管理、多人遊戲、社群網路服務等功能。Steam 平台使用者能安裝並自動更新遊戲，還可以進行遊戲紀錄雲端儲存、遊戲內建的語音系統、遊戲親友共享及遊戲成就系統等功能。Steam 平台提供數萬款遊戲，同時也擁有龐大的玩家社群，截至 2021 年，在 Steam 上架的總遊戲數量已超過 50,000 款，且已約有 10 億註冊帳戶，同時在線的活躍玩家高達 2,500 萬(簡育詮, 2021)。

而在電子遊戲的眾多遊戲類型中，角色扮演遊戲(Role-Playing Game, RPG)市場佔有全球遊戲的收益的很大部分。角色扮演遊戲允許玩家扮演虛擬角色，以解決任務、與其他遊戲虛擬角色互動、擊敗敵人並通過獲得各種技能和裝備，讓角色變得更強大(Grouling, 2010)。目前角色扮演遊戲多以中世紀奇幻主題為主，但也有現代和科幻角色扮演遊戲。但廣義而言，玩家有扮演虛擬角色的遊戲，均可被視為 RPG 類型遊戲。

但對於遊戲相關的研究，過去較常使用問卷調查玩家對電子遊戲的看法和態度(Phan, Keebler & Chaparro, 2016)。然而，問卷大多使用量表系統，使用者無法通過量表來表達意見，遊戲開發團隊可能無法真正了解玩家的想法。而文字探勘技術(Text Mining)可以分析遊戲玩家的評論，從評論中獲取最直接的玩家想法。基於上述原因，本研究決定使用文字探勘技術，通過分析遊戲評論來探索玩家的感知。

1.2 研究動機

使用者生成內容(User Generated Content, UGC)對於其他潛在使用者或管理者非常重要。使用者對於不同商品與服務的意見，不僅能顯現使用者的個性特徵，對公司未來的決策及其他使用者的選擇也會造成強烈的影響 (Giachanou & Crestani, 2016)。目前許多消費者都在減少衝動性購買行為，意味著使用者評論的重要性愈來愈高，在電子遊戲產業中此現象尤為明顯。但由於各款遊戲的線上評論所提供的文本內容數量眾多，且評論內容多半凌亂複雜，且多不具結構性。

以 Steam 平台為例，玩家評論中的主題可能與遊戲屬性，包括角色、遊戲畫面、遊戲劇情、挑戰等有一定關係，這些主題可以視為玩家的偏好特徵。而為了找到這些遊戲評論的主題，本研究使用常見的方法便是將文本套用於主題模型。對於遊戲開發團隊來說，評論應是開發團隊對其次世代產品開發最有價值的參考資料。以下問題為本研究的主要探討議題：

- (1) 玩家心目中喜愛的遊戲需具備何種條件？
- (2) 未來開發團隊應該朝什麼方向改進其遊戲品質，以提升遊戲玩家評價？

1.3 研究架構

本研究將確認所要探討的研究主題，了解其研究相關環境及背景，進行相關文獻的蒐集與分析，以確定研究的範圍與架構，第一章為本研究的背景簡介，第二章探討相關文獻，第三章描述本研究所採用的內容分析方法，第四章進行資料分析與討論，第五章總結本研究並提出未來研究方向的建議。圖 1.1 為本研究之架構圖。

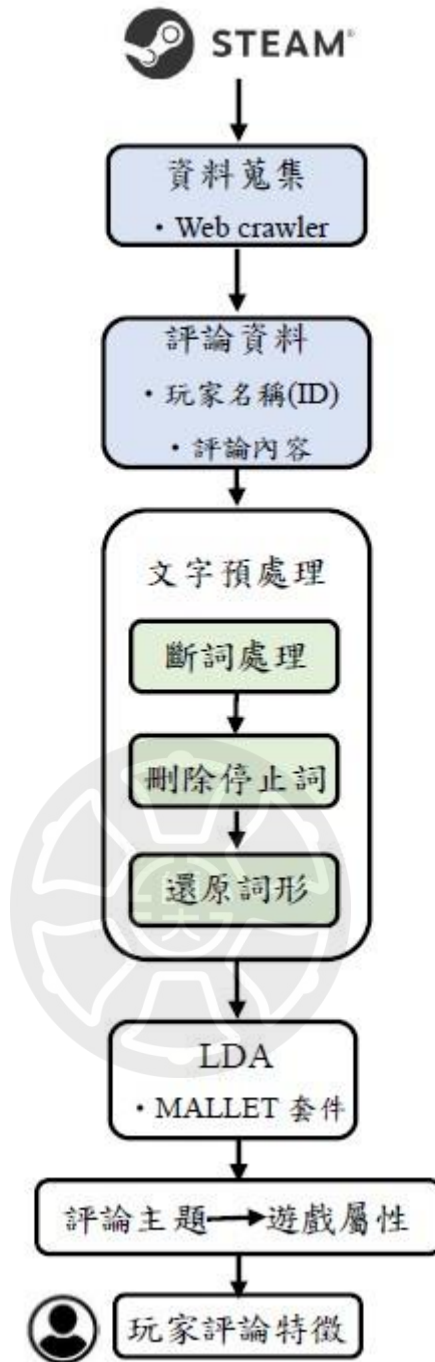


圖 1.1 研究架構圖
資料來源：本研究整理

第二章 文獻探討

2.1 使用者生成內容與評論的應用

使用者生成內容(簡稱UGC)，指由所有使用者在網路上創作公開原創內容並將其分享(Daugherty, Eastin, & Bright, 2008)。使用者自願提供數據、訊息或媒體，然後以實用的或有趣的方式呈現於他人。近年來，使用者生成內容在社群媒體上的應用迅速增長，原因在於獲取內容方便以及其公開的特性(Krumm, Davies & Narayanaswami, 2008)。

目前擁有大量使用者的平台如Steam、YouTube、Facebook、Wikipedia及Instagram等等社群平台都屬於使用者生成內容的範疇，使用者可以隨時在平台上創作屬於自己的內容，而這些由使用者發布的不同形式的內容可能包括影音、文字內容等(Clever, Kirchner, Schray & Schulte, 2009)。以文本式評論來說，消費者可以在Amazon線上商店上對買過的產品撰寫使用心得，旅行者則可以在TripAdvisor.com上對旅遊規畫進行評論，觀眾能在Rotten Tomatoes網站上發表自己的影評(Strååt & Verhagen, 2017)。而研究表明使用者生成內容也廣泛應用於遊戲產業，諸如像是Steam、Metacritic等平台都有提供玩家評論功能。

從過去的研究發現，遊戲的實際評分與玩家體驗之間的關係相對較小，但是玩家評論內容可以概括大部分玩家的遊戲體驗並形成主觀評價，代表評價可以是玩家感受的直接體現，故研究建議可以從使用者的評論內容進行文本分析和探勘(Johnson, Watling & Gardner, 2014)。像是過去的研究評估了使用者評論在電子遊戲產業中的影響(Sherrick & Schmierbach, 2016)。此外遊戲評論也包括玩家針對遊戲的認知和情感結果(Kochler, Arnold, Greenhalgh & Boltz, 2017)。總結來說，對於大部分玩家而言，遊戲評論將會是評估是否要購買該款遊戲的重要依據(Johnson et al., 2014)。

2.2 主題模型(Topic Model)與隱含狄利克雷分布(Latent Dirichlet Allocation)

主題模型是在自然語言處理(Natural Language Processing, 簡稱 NLP)與機器學習(Machine learning)等領域, 用以處理文本中發現的抽象主題的統計模型。而最早的主題模型由 Papadimitriou, Raghavan, Tamaki 與 Vempala 在 1998 年所提出。在自然語言處理及機器學習中, 許多主題模型都透過字詞的頻率來計算字詞的特徵, 特徵值越高代表該字詞越接近該篇文章主題。而有關主題模型的應用相當多, 這些研究多半是將大量文本進行集中加以學習, 找出在文本內所隱藏的主題, 接著再將其他文本套入到已經訓練好的模型之中, 並計算出新文本主題的分布以進行分析。

學者 Blei, Ng 與 Jordan 於 2003 年提出隱含狄利克雷分佈(Latent Dirichlet Allocation, LDA), 為一種主題建模演算法(Topic modeling), 以非監督式學習的方式探索大型文件集合或語料庫(Corpus)中的潛在主題, 而潛在主題由不同隨機分布的字詞組成, 且相同的字詞可能同時出現在不同的主題, 圖 2.1 顯示 LDA 中主題-字詞-文件之間的關聯。

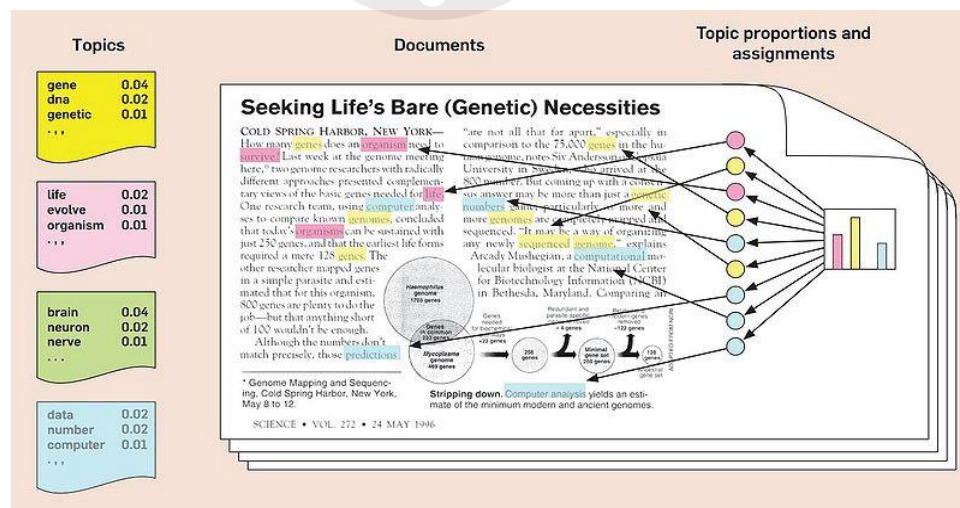


圖 2.1、主題模型示意圖

資料來源：Blei, D. M. (2003)

隱含狄利克雷分布是最常見的機率主題模型之一，是一種有用的文本挖掘技術，並用於分析文本格式的大數據(Tirunillai & Tellis, 2014)。LDA 通過分析主題分佈和單詞分佈，確定文本的元素。文本內主題和字詞的分佈都具有 Dirichlet 分布(Hong & Davison, 2010)。

本研究搜索過去研究中，使用不同種類文本與運用主題模型作特徵分析的相關研究並與本研究進行比較，詳細列表請參考表 2.1。Tong 與 Zhang (2016)運用 LDA 分析，先以維基百科詞條建立主題模型，再分析 Twitter 使用者的推文，並找出使用者的興趣。而 Kim 與 Kim (2022)則使用一種 ElasticSearch 分類方法和 LDA 模型來提取學術論文的特徵。提取關鍵詞搜索結果的主題，對論文搜索結果進行分類。但由於非結構化的性質，研究人員較難使用探索性因素分析(Exploratory factor analysis，簡稱 EFA)來分析使用者生成的內容。此外，研究人員無法在短時間內分析大量訊息的主題(Liu, Burns & Hou, 2017)。基於這些問題，本研究使用 Python 的 LDA 相關第三方套件來擷取遊戲評論中的潛在主題。

表 2.1、主題模型研究相關整理

研究者	文本來源	研究目標
Tong & Zhang (2016)	Wikipedia, Tweet	以維基百科詞條建立主題模型，分析 Twitter 使用者的推文，找出使用者的興趣
Kim & Kim (2022)	Google Scholar	應用 ElasticSearch 分類方法和 LDA 模型來提取學術論文的特徵。提取關鍵詞搜索結果的主題，對論文搜索結果進行分類。
本研究	Steam	從 RPG 的遊戲評論，找出多數玩家關注的主題，並對主題內的字詞以遊戲屬性做分類。

資料來源：本研究整理

2.3 遊戲屬性分類

對於遊戲屬性分類，不同研究會有分出不同屬性，導致各種屬性重複性過高的問題。根據 Bedwell, Pavlas, Heyne, Lazzara 與 Salas(2012)的研究提出一種統一衡量遊戲的屬性。Bedwell 提出的分類法共九項屬性，分別是(1)動作語言、(2)評估、(3)衝突/挑戰、(4)控制，(5)環境、(6)遊戲劇情、(7)角色互動、(8)沉浸感及(9)規則/目標，如表 2.1 所示。

首先“動作語言”是指玩家向遊戲系統下達命令以控制遊戲中角色或事件的方式。例如按下鍵盤上的“W”、“A”、“S”、“D”按鍵以使遊戲中的角色前進後退或是左右移動(Owen, 2004)。“評估”指在遊戲過程中對玩家的反饋，包括遊戲成就的衡量和玩家的成長(Chen & Michael, 2005；Owen, 2004)。例如，在遊戲內完成某些事件或任務，就能獲得成就。遊戲結束後，玩家獲得的成就比分將顯示出來。“衝突/挑戰”是指遊戲內謎題的形式和難度，其中包括適應性、挑戰性、衝突性(Gentile & Gentile, 2008)。“控制”則是指玩家可以在遊戲中影響的程度，包括角色控制或與遊戲事件的互動。“環境”是指玩家所處遊戲環境，包括遊戲物件的配置或是模擬風景的呈現等(Owen, 2004)。“遊戲劇情”是指遊戲世界和故事，其中包含幻想（與現實世界不同的事物）和神秘感（已知與未知之間的鴻溝）(Garris, Ahlers, & Driskell, 2002；Owen, 2004)。“互動性”是指遊戲中人與人之間的任何接觸(Crawford, 1984)。而人際互動可以包括彼此依存和競爭行為。例如某些多人連線遊戲提供即時通訊或語音聊天，幫方便玩家們彼此溝通。“沉浸感”是指玩家對遊戲劇情的感知和情感，包括表現力及感覺刺激(Owen, 2004)。“規則/目標”包括規則與目標的清晰性。規則是指遊戲的限制，而目標是指玩家採取行動的意圖(Owen, 2004)。

透過 Bedwell 的屬性分類法能清楚描述各種遊戲元素，使研究人員可以對評論主題進行分類，幫助研究人員確定玩家經常討論遊戲主題的內容。

表 2.2、遊戲屬性說明

屬性名稱 (譯名)	屬性說明
Action Language (動作語言)	玩家向遊戲系統下達指令以控制遊戲中事物的方式
Assessment (評估)	遊戲中玩家的反饋
Conflict/Challenge (衝突/挑戰)	遊戲中問題的形式和難度
Control (控制)	玩家可以在遊戲中的影響程度
Environment (環境)	遊戲角色所在的環境
Game Fiction (遊戲劇情)	遊戲的世界與故事
Human Interaction (互動性)	遊戲中角色與角色間的互動
Immersion (沉浸感)	玩家對遊戲劇情的感知和情感
Rules/Goals (規則/目標)	遊戲規則和目標資訊的清晰性

資料來源：本研究整理

第三章 研究方法

本研究所提出的內容分析方法為先選擇目標遊戲，再進行線上評論的資料蒐集並清洗，以取得研究所需資料。接著使用主題模型找出遊戲評論的主題，並將主題套用在 Bedwell 所定義的遊戲屬性中，從而找出多數玩家對何項的遊戲屬性較有感知。

3.1 資料蒐集

本研究使用 Python 的第三方套件 steamreviews，撰寫網路爬蟲程式 (Web Crawler)，從遊戲社群網站 Steam 平台蒐集三款熱門的角色扮演遊戲 (Role-Playing Game，簡稱 RPG)，其中包含平台所提供之玩家的評論，並儲存成 CSV 檔。通過評論內容，運用 LDA 主題模型找到更多有關玩家評論特徵及遊戲屬性的線索。

表 3.1、三款 RPG 遊戲所對應之評論數量(擷取評論時間到 2021.5 為止)

遊戲名稱	評論數
電馭叛客 2077	95,471
巫師 3：狂獵	96,898
隻狼：暗影雙死	50,506

資料來源：本研究整理

評論擷取程式皆依照下列條件進行：(1)針對三款 RPG 的“英文”評論進行蒐集。由於 Steam 提供多種語言評論，如英文、中文、日文等，而英文是世界最廣泛使用的語言；(2)Steam 每款遊戲的評論數量均有所不同，多款熱門遊戲的評論數量多達上萬則，較為冷門的評論數可能只有數百則。為求一致性，本研究設定三款 RPG 遊戲所擷取的評論數量共 242,875 則評論，

三款遊戲分別的評論數量如表 3.1 所示；(3)在角色扮演遊戲中同時也包含許多不同的遊戲機制，而本研究所選擇的三款 RPG 遊戲皆有不同內容與故事背景。而這三款 RPG 的遊戲評論數最多，為該類別的熱門遊戲並同時具有代表性，故選擇以此三款遊戲作為 RPG 的研究目標。

評論資料包括：(1)玩家名稱(ID)及(2)評論內容，如圖 3.1 所示，其中玩家名稱代表該則評論為一位玩家所評論，而評論內容代表該玩家對於遊戲體驗的直接意見。



圖 3.1、Steam 評論之使用者介面
資料來源：Steam 網站

3.2 文字預處理

為了獲取文字資料中的關鍵特徵，事前的文字預處理是必要的步驟。在使用 steamreviews 套件下載評論資料後，由於評論內容中會混雜一些非英文或是代碼(如 u\2019 為撇號代碼，原文 I don't 會辨識為 I donu\2019t)，故需要進行初步整理，並將所有遊戲評論的字詞整理成 CSV 檔，以利後續進行分析。在經過處理後的評論資料如圖 3.2 所示。

	A	B	C	D
1	recommendsteamid		review	voted_up
2	"96280169"	"76561199007484140"	"I never write reviews and I'll keep this short. I was excited when	FALSE
3	"96280086"	"76561198248627448"	"This game is has an amazing story if you can get past a few bug	TRUE
4	"96279981"	"76561198003368135"	"Yes, but with a caveat: Cyberpunk 2077 is uneven and has fund	TRUE
5	"96277126"	"76561198022582231"	"A fantastic game with an excellent story, great combat mechanic	TRUE
6	"96277109"	"76561198117317257"	"I wanted so much to love this game. I followed the developmen	TRUE
7	"96274729"	"76561198856481190"	"Play as a male, Remove your pants ;) Thank youuuuuuu."	TRUE
8	"96273019"	"76561198346184095"	"7/10 i love this game"	FALSE
9	"96270337"	"76561199072826141"	"its just perfect"	TRUE
10	"96270150"	"76561198241573268"	"Cyberpunk 2077 is a pretty good game in my opinion, on PC, o	TRUE
11	"96269389"	"76561198105268811"	"awful game"	FALSE
12	"96268893"	"76561197994543968"	"Enjoyed from start to finish, ignored the bugs, got the feels. Will	TRUE
13	"96268878"	"76561198070803135"	"The game has come a long way. Only very few times do I get a	TRUE
14	"96268548"	"76561198049519182"	"Not as advertised, but fun nevertheless."	TRUE
15	"96268485"	"76561198165691144"	"If your pc can handle it."	TRUE

圖 3.2、遊戲評論資料示意圖

資料來源：本研究整理

接著本研究使用 Python 的自然語言處理工具(Natural Language Tool Kit, NLTK)，將評論內容進行文字預處理，將過程分為三階段：(1)斷詞處理(Tokenize)、(2)刪除停止詞(Stop words)及(3)還原詞形(Lemmatization)。

(1)斷詞處理在自然語言的研究上是重要的步驟，主要就是將原始句子依據英文結束符號作為斷詞斷句的判斷，例如句號(.)、問號(?)、驚嘆號(!)、連接號(-)等，再根據句子中的空格將英文單詞分隔開，取得獨立的字詞。

(2)刪除停止詞會將所有字詞進行篩選處理，將無意義的字詞(即停止詞)去除後，使語料庫只含有關鍵的字詞，停止詞列表由表 3.2 所示。經過斷詞處理後，每款遊戲評論文本已經可以由字詞的集合所表示，但由於字詞中可能還存在一些對研究沒有相關意義的字詞、特殊符號等，若保留這些無意義的字詞，將會變成之後訓練模型的累贅，因此本研究使用 NLTK 內建之停止詞進行去除。

(3)還原詞形主要是將字詞轉變為字本身基本形式的過程。還原詞形與提取詞幹(Stemming)皆是將字詞的不同型態歸一來降低文本複雜度的常見方式，兩者之間的區別是詞形還原會考慮上下文並將字詞轉換為有意義的基本形式，而詞幹提取則直接移除字尾，可能會導致錯誤的涵義及拼寫錯誤。

以圖 3.2 中的第一條評論為例，該評論原始內容為"I never write reviews and I'll keep this short. I was excited when it came out. I experienced some glitches, that's not the problem"，在經過文字預處理後會分割為以下字詞：'never', 'write', 'reviews', 'keep', 'short', 'excited', 'come', 'experienced', 'glitches', 'problem'，去除停止詞並保留有意義的字詞，最後將字詞還原成原來樣貌。文字預處理的完整程式碼請參考附錄一。

表 3.2、停止詞列表

I	me	my	myself	we	us	our	ourselves
you	your	yourself	he	him	his	himself	she
her	herself	it	its	itself	they	them	their
themselves	what	who	when	where	why	which	how
this	that	above	below	under	do	does	did
was	were	be	been	being	as	until	while
of	to	about	against	into	out	in	at
from	further	here	there	up	down	with	...

資料來源：NLTK Stop Words(<https://www.tutorialexample.com/list-all-english-stop-words-in-nltk-nltk-tutorial/>)

3.3 主題模型應用

本研究以歸納出主題的意義，夠透過主題的結果表達各遊戲評論主題之間的關聯性與相似度，進而找出文本中隱含的訊息。使用主題模型之前，需要透過套件計算出三款遊戲個別的最適主題數量，如表 3.3，找到最適主題的數量就能夠有挖掘出最合適的主題內容。評斷出最適主題數的方法為一致性分數(Coherence Score)，一致性分數代表在該主題中各字詞的相關性分數。當分數越高，表示在單一主題內所包含的字詞當中的越有相關

性。運用主題模型得到的結果會利用 Sievert & Shirley (2014)所開發的一個視覺化套件 LDAvis，繪出個別遊戲在各主題數量下的一致性分數。結果如表 3.3 所示。

表 3.3、各遊戲最適主題數量

遊戲名稱	一致性分數結果	主題數量																
電馭叛客 2077	<table border="1"> <caption>Data for Cyberpunk 2077</caption> <thead> <tr> <th>Num Topics</th> <th>Coherence score</th> </tr> </thead> <tbody> <tr><td>5</td><td>0.415</td></tr> <tr><td>10</td><td>0.455</td></tr> <tr><td>15</td><td>0.480</td></tr> <tr><td>20</td><td>0.475</td></tr> <tr><td>25</td><td>0.480</td></tr> <tr><td>30</td><td>0.470</td></tr> <tr><td>35</td><td>0.475</td></tr> </tbody> </table>	Num Topics	Coherence score	5	0.415	10	0.455	15	0.480	20	0.475	25	0.480	30	0.470	35	0.475	14
Num Topics	Coherence score																	
5	0.415																	
10	0.455																	
15	0.480																	
20	0.475																	
25	0.480																	
30	0.470																	
35	0.475																	
巫師 3： 狂獵	<table border="1"> <caption>Data for The Witcher 3: Wild Hunt</caption> <thead> <tr> <th>Num Topics</th> <th>Coherence score</th> </tr> </thead> <tbody> <tr><td>5</td><td>0.500</td></tr> <tr><td>10</td><td>0.528</td></tr> <tr><td>15</td><td>0.538</td></tr> <tr><td>20</td><td>0.532</td></tr> <tr><td>25</td><td>0.528</td></tr> <tr><td>30</td><td>0.535</td></tr> <tr><td>35</td><td>0.538</td></tr> </tbody> </table>	Num Topics	Coherence score	5	0.500	10	0.528	15	0.538	20	0.532	25	0.528	30	0.535	35	0.538	14
Num Topics	Coherence score																	
5	0.500																	
10	0.528																	
15	0.538																	
20	0.532																	
25	0.528																	
30	0.535																	
35	0.538																	
隻狼：暗 影雙死	<table border="1"> <caption>Data for Sekiro: Shadows Die Twice</caption> <thead> <tr> <th>Num Topics</th> <th>Coherence score</th> </tr> </thead> <tbody> <tr><td>5</td><td>0.400</td></tr> <tr><td>10</td><td>0.445</td></tr> <tr><td>15</td><td>0.468</td></tr> <tr><td>20</td><td>0.470</td></tr> <tr><td>25</td><td>0.465</td></tr> <tr><td>30</td><td>0.458</td></tr> <tr><td>35</td><td>0.465</td></tr> </tbody> </table>	Num Topics	Coherence score	5	0.400	10	0.445	15	0.468	20	0.470	25	0.465	30	0.458	35	0.465	20
Num Topics	Coherence score																	
5	0.400																	
10	0.445																	
15	0.468																	
20	0.470																	
25	0.465																	
30	0.458																	
35	0.465																	

資料來源：本研究整理

經過 LDAvis 套件的計算及結果後，分別得出三款遊戲評論的一致性分數，分數最高即代表在該遊戲選擇的主題數量下，其主題之中的字詞最有相關性，而該主題數量及代表遊戲評論的最適主題數量。而三款遊戲獲得最高一致性分數的最適主題數量分別為(1)電馭叛客 2077，14 個主題(2)巫師 3：狂獵，14 個主題(3)隻狼：暗影雙死，20 個主題。

在得出最適主題數量後，即可開始建構主題模型。從個別的游戏評論資料，創建主題建模所需的詞典與語料庫(Corpus)。詞典與語料庫資料的資料來源皆為從 Steam 平台使用爬蟲程式擷取的遊戲評論文本。而三款遊戲評論的輸出結果是經過 LDA 分析後的個別主題列表，每一個主題都還有一個相關的熱門詞列表。同時每個主題都有若干個百分比來表示其在文檔中的分佈概率，構成該主題的詞表中包含了表示每個字詞屬於該主題的概率的係數。計算一致性分數的完整程式碼請參考附錄二。



第四章 研究結果

4.1 研究資料

本研究根據第 3.1 節所提出的準則，運用網路爬蟲擷取 Steam 三款熱門 RPG 的評論，經過第 3.2 節的文字預處理後剩下共 242,875 則評論。本研究運行 LDA 主要使用 Python 第三方套件 MALLET，該套件是基於 Java 語言的套件，用於對文本進行自然語言處理、文檔分類、主題建模、訊息擷取和其他機器學習程序。表 4.1 為執行 LDA 所設定的參數，完整程式碼可參考附錄三。

表 4.1、LDA 設定參數

LDA 參數	參數說明
corpus	由評論經過斷詞成為單一字詞所組成的語料庫
id2word	語料庫內所有字詞的編號
num_topic	決定分析的主題數量

資料來源：本研究整理

4.2 結果評估

4.2.1 電馭叛客 2077

電馭叛客 2077 為波蘭遊戲開發商 CD Projekt Red 所開發，於 2020 年 12 月在 Steam 等遊戲平台上架發售。電馭叛客 2077 為一款第一人稱射擊角色扮演遊戲，由玩家扮演客製化的虛擬角色，並探索一個賽博龐克風格 (Cyberpunk) 的遊戲世界。

表 4.1 為該遊戲評論擁有最廣分布的前三個主題。主題內字詞後方的百分比則代表該字詞在主題內的權重。結果如表 4.1 所示，其中有關於規則/目標(Rules/Goals)主題(33.3%)；控制(Control)主題(17.9%)；及環境(Environment)主題(9.2%)有最多的討論。電馭叛客 2077 遊戲的玩家在 Steam 上分享對該遊戲的看法。玩家對遊戲的規則及目標(Rules/Goals)由很

高的討論度，主題內字詞如 AI、System、Mechanic、Gameplay、Content 代表玩家在體驗該遊戲時，相當注重遊戲本身的核心玩法。其次討論的主題是控制(Control)，該遊戲的故事劇情會因為玩家本身的選擇而有些許不同，故在主題內的字詞多半提到如 Choice、Option、End、Dialogue 等字詞。第三個討論主題是環境(Environment)，該主題關鍵詞如 Open、World 等字詞，同時提到其他遊戲(Gta, Fallout, Witcher)，也同樣屬於開放世界 RPG，但遊戲場景與該遊戲有所不同，代表玩家對遊戲的開放世界元素相當重視。

表 4.2、電馭叛客 2077 之分布佔比前三個主題(完整主題列表請參考附錄四)

主題編號	分布	主題中前十個字詞	屬性標籤
T6	0.333	feel(.058); ai(.023); lack(.018); system(.018); feature(.018); miss(.016); bad(.016); mechanic(.016); gameplay(.014); content(.013)	Rules/Goals
T13	0.179	character(.067); choice(.032); end(.028); feel(.023); option(.023); change(.019); make(.019); dialogue(.017); matter(.016); choose(.013)	Control
T1	0.092	world(.123); open(.052); game(.051); rpg(.047); cyberpunk(.041); expect(.023), gta(.023); fallout(.018); witcher(.016); action(.012)	Environment

資料來源：本研究整理

4.2.2 巫師 3：狂獵

巫師 3：狂獵同為 CD Projekt Red 所開發的第三人稱開放式的角色扮演遊戲，該遊戲為巫師系列第三款作品。故事背景為中古奇幻世界。主要遊戲方式為玩家使用武器和魔法戰鬥，完成任務以獲取經驗值和金錢，用於

提升的能力和購買裝備。

表 4.2 為該遊戲評論擁有最廣分布的前三個主題。在所有評論中，對於規則/目標(Rules/Goals)主題(44.4%)；及沉浸感(Immersion)主題(11.1%)有最多討論。說明玩家對於該遊戲的規則及目標(Rules/Goals)有很高的關注，在編號 T0 與 T3 主題中的大部分字詞都代表該遊戲的核心玩法，如 Level 代表 RPG 的等級機制，Fight、Sword 代表遊戲以劍戰鬥為主，Quest、Side、Mission 等詞代表遊戲的主要及支線任務等。其次討論的主題是沉浸感(Immersion)，該主題的字詞 Beautiful、Fantasy、Immersive 等，代表許多玩家使用正面的字詞來稱讚該遊戲對於故事、圖形所帶給玩家等的沉浸感。Soundtrack 則表示玩家對遊戲的配樂很重視，而配樂也帶給玩家更好的遊戲體驗，讓玩家更沉浸在遊戲世界。

表 4.3、巫師 3：狂獵之分布佔比前三個主題(完整主題列表請參考附錄五)

主題編號	分布	主題中前十個字	屬性標籤
T0	0.351	level(.028); fight(.020); system(.018); enemy(.017); combat(.014); difficulty(.014); horse(.012); sword(.012); high(.011); potion(.011)	Rules/Goals
T12	0.111	world(.177); open(.070); character(.064); beautiful(.027); detail(.020); immersive(.019); fantasy(.016); engage(.016); combat(.016); soundtrack(.014)	Immersion
T3	0.093	quest(.157); story(.085); side(.079); main(.062); gwent(.039); interesting(.026); explore(.022); find(.022); mission(.021); card(.020)	Rules/Goals

資料來源：本研究整理

4.2.3 隻狼：暗影雙死

隻狼：暗影雙死為日本遊戲開發商 From Software 於 2019 年所開發的第三人稱動作角色扮演遊戲。相較於一般 RPG，遊戲少了角色自定義功能。該遊戲比起探索及蒐集要素，該遊戲更加著重於動作要素，難度比起一般 RPG 高，故遊戲社群稱此類高難度動作的 RPG “類魂” 遊戲。

表 4.3 為該遊戲評論擁有最廣分布的前三個主題，在所有評論中，規則/目標(Rules/Goals)主題(20.6%)；衝突/挑戰(Conflict/Challenge)主題(10.6%)；及動作語言(Action Language)主題(6.4%)最多討論。

表 4.4、隻狼：暗影雙死之分布佔比前三個主題(完整主題列表請參考附錄六)

主題編號	分布	主題中前十個字	屬性標籤
T9	0.206	enemy(.089); attack(.087); deflect(.034); posture(.031); block(.026); damage(.022); health(.022); dodge(.020); hit(.020); counter(.018)	Rules/Goals
T5	0.106	hard(.164); love(.0145); awesome(.041); hate(.041); defeat(.036); cool(.032); super(.028); hesitation(.022); man(.020); fuck(.016);	Conflict/Challenge
T17	0.064	enemy(.054); kill(.030); camera(.022); hit(.021); run(.015); bad(.014); move(.014); stealth(.013); back(.012); animation(.011);	Action Language

資料來源：本研究整理

根據表 4.3，隻狼：暗影雙死的玩家在 Steam 上分享對該遊戲的看法。規則及目標(Rules/Goals)屬性佔該遊戲評論資論資料最廣分布，代表核心玩法為玩家最重視的議題，主題內字詞如 Attack、Deflect、Posture、Block、Dodge、Counter 代表遊戲以招架敵人攻擊並反擊為核心玩法，這與前兩款 RPG 有所不同。其次討論的主題是衝突與挑戰(Conflict/Challenge)，該主題 Hard、Defeat、Hate 代表該遊戲對於玩家有一定難度，而 Love、Awesome 等字詞則表示玩家可能享受該遊戲的挑戰。最後玩家對於動作語言(Action Language)相當重視。該遊戲的其中一個特色就是動作要素，主題中的 Move、Back 代表遊戲角色的動作，而 Camera 表示遊戲由滑鼠所操控的視角，以動作要素為主的遊戲來說，視角靈敏度對於玩家體驗有很大影響。

4.3 綜合討論

表 4.4 為各項研究結果之統整，從表得知，以 RPG 為例，規則與目標為玩家共同關注的議題。但三款遊戲個別相異屬性對應三款遊戲不同的要素。總結來說，同為 RPG 類型但不同遊玩要素也會對玩家評論的遊戲屬性有所不同，因此遊戲開發商在選擇研究對象時要特別注意與待開發遊戲的遊玩要素是否一致。

表 4.5、三款遊戲之屬性標籤

遊戲名稱	屬性標籤
電馭叛客 2077	規則/目標(Rules/Goals)、控制(Control)、環境(Environment)
巫師 3：狂獵	規則/目標(Rules/Goals)、沉浸感(Immersion)
隻狼：暗影雙死	規則/目標(Rules/Goals)、衝突/挑戰(Conflict/Challenge)、動作語言(Action Language)

資料來源：本研究整理

第五章 結論與建議

5.1 結論

以往對遊戲評論的研究主要集中在分析專業評論和銷售之間的關係，或者說專業評論的能力影響玩家的感及和態度(Johnson et al., 2014; Sherrick & Schmierbach, 2016)，較少研究電子遊戲玩家評論的看法和態度。本研究應用主題模型並有效地從電子遊戲平台 Steam 推導遊戲評論的主題。通過對大多數玩家的遊戲評論進行分析，發現玩家真正關注的遊戲議題。研究表明評論資料能為研究人員探索玩家的感知及態度提供了大量且多樣的數據來源(Johnson et al., 2014; Koehler et al., 2017; Strååt & Verhagen, 2017)，並對遊戲評論進行了實際驗證。總結來說，本研究將評論大數據轉化為可分析的資料，並回答與遊戲開發相關的研究問題，(1)以本研究為例，RPG 玩家最重視就是遊戲規則以及遊戲目標，但不同核心玩法就會讓玩家對遊戲有不同偏好，故可以透過分析評論的主題來尋找相異性。(2)即使是不同種類的 RPG，開發人員也需要從遊戲評論資料中的主題找出相異的遊戲屬性，若僅從遊戲共同的屬性來真改善可能會無法正確了解玩家的偏好，故應著重於相異主題才能提供遊戲後續的更新及續作的改善方向。

5.2 研究限制

本研究運用爬蟲程式擷取遊戲評論資料進行分析，並找出評論內容的主題，然而本研究仍受限於以下限制：

- (1) 擷取遊戲評論資料時，設定為三種類別之熱門遊戲的英文玩家評論，共 242,875 筆資料。在玩家評論內容過少的情況下可能無法準確擷取玩家特徵，為了在後續能更準確地分析玩家偏好及相似玩家，本研究在評論篩選時，只選擇英文之評論資料，但僅有英文評論可能會有代表性不足的問題。資料豐富度不夠就可能影響到分析

玩家偏好的主題。

- (2) 本研究將所擷取之玩家評論資料進行主題分析，其內容越接近的主題即為該玩家的偏好。Steam 平台提供的評論篩選功能，可以檢視該玩家評論時，是否標示為正面或負面評論。本研究並未對評論進行篩選，原因在於本研究所選擇的遊戲負面評論不足，可能無法生成有效的結果，故將正面及負面評論資料一同分析。
- (3) LDA 不太適合使用於簡短的文本上。遊戲評論長短不一，若只有幾個單詞的簡短評論，可能會發生的評論稀疏問題，而無法用得出的主題來正確的推斷遊戲屬性，可能會造成過度解讀。

5.3 未來研究方向

第一，本研究中遊戲評論僅來自 Steam 平台，而評論來源的不足則可能忽略掉更多玩家對不同遊戲的看法。對於評論的多樣性，未來的研究人員可以擴展到其他電子遊戲平台，例如像是 Metacritic、GameSpot 等，不同平台使用者可能也會對同一款遊戲有不同的偏好，故能擷取來自其他玩家的評論使研究資料更為豐富。

其次，LDA 需要更有效和準確的方法來驗證輸出。Grimmer 與 Stewart (2013)提到機器學習的文本分析應該與更有可信度的驗證方法一起使用。而 Maier 等人(2018)提出了一種應用的方法來提高 LDA 主題模型結果的可靠性和有效性。例如，適當的文本集合預處理、模型參數的適當選擇、主題數量等。

第三，未來也可以對其他種類遊戲的評論資料進行研究，找出該類型遊戲玩家的偏好並進行比較，像冒險遊戲、策略遊戲、格鬥遊戲等諸多分類(Poole, 2011)可能都會產生不同主題。本研究所選擇的角色扮演遊戲有相同的主題，但在不同核心玩法下主題便有所不同。各類型遊戲都會有不同的核心玩法，研究人員必須了解研究的遊戲屬於何種類型，甚至是類型下

不同核心玩法的子類型，才能獲取正確的主題。

最後，Steam 雖然提供正面與負面評論的分類，但並非所有評論者都會正確標示為正面評論或負面評論，評論的內容才能代表玩家的想法。未來能結合文本情感分析，運用所分析出的主題作為特徵字，並找出評論中的情緒詞彙。舉例來說，一則評論內容為 "This game is full of glitches, that's a huge the problem"，其中 "game" 為主題，"glitches"、"problem" 代表負面情緒字詞，表示該玩家對於遊戲漏洞相當不滿意。運用此方法分析玩家對主題正面及負面的情感，幫助研究人員更了解玩家真實的想法。。



參考文獻

- Bedwell, W., Pavlas, D., Heyne, K., Lazzara, E. H., & Salas, E. (2012). Toward a taxonomy linking game attributes to learning: An empirical study. *Simulation & Gaming, 43*(6), 729-760.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Learning research, 3*, 993-1022.
- Chen, S., & Michael, D. (2005). Proof of learning: Assessment in serious games. Gamasutra. Retrieved April 27, 2021 from <https://www.gamedeveloper.com/design/proof-of-learning-assessment-in-serious-games>
- Clever, N., Kirchner, A., Schray, D., & Schulte, M. (2009). User-generated content. Informally published manuscript, Institute for Economic Computer Science, Münster, Germany.
- Crawford, C. (1984). *The art of computer game design*. Berkeley, Calif: Osborne/McGraw-Hill.
- Daugherty, T., Eastin, M. S., & Bright, L. (2008). Exploring consumer motivations for creating user-generated content. *Journal of interactive advertising, 8*(2), 16-25.
- Facebook Gaming (2021). Gaming marketing insight for 2021. Retrieved May 15, 2021 from <https://www.facebook.com/business/news/insights/what-unprecedented-growth-means-for-the-future-of-the-gaming-industry>
- Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*(4), 441-467.
- Gentile, D. A., & Gentile, J. R. (2008). Violent video games as exemplary teachers: A conceptual analysis. *Journal of Youth Adolescence, 37*(2), 127-141.

- Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys (CSUR)*, 49(2), 28.
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3), 267-297.
- Cover, J. G. (2010). *The creation of narrative in tabletop role-playing games*. Jefferson, N.C: McFarland & Co.
- Hong, L., & Davison, B. D. (2010). Empirical study of topic modeling in twitter. *Proceedings of the first workshop on social media analytics*, 80-88.
- Johnson, D., Watling, C. N., Gardner, J., & Nacke, L. E. (2014). The edge of glory: The relationship between metacritic scores and player experience. *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play*, 141-150.
- Kim, M., & Kim, D. (2022). A Suggestion on the LDA-Based Topic Modeling Technique Based on ElasticSearch for Indexing Academic Research Results. *Applied Sciences*. 12, 3118.
- Koehler, M. J., Arnold, B., Greenhalgh, S. P., & Boltz, L. O. (2017). A taxonomy approach to studying how gamers review games. *Simulation & Gaming*, 48(3), 363-380.
- Krumm, J., Davies, N., & Narayanaswami, C. (2008). User-Generated Content, *IEEE Pervasive Computing*, 7(4), 10-11.
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science*, 343(6176), 1203-1205.
- Lin, D., Bezemer, C., & Hassan A.E. (2017). Studying the urgent updates of popular games on the Steam platform. *Empirical Software Engineering*, 22(4), 2095-2126.

- Lin, D., Bezemer, C., Zou, Y. & Hassan, A. E. (2019). An Empirical Study of Game Reviews on the Steam Platform. *Empirical Software Engineering*, 24, 170-207
- Lin, M., Lucas Jr, H. C., & Shmueli, G. (2013). Too big to fail: Large samples and the p-value problem. *Information Systems Research*, 24(4), 906-917
- Liu, X., Burns, A. C., & Hou, Y. (2017). An investigation of brand-related user generated content on Twitter. *Journal of Advertising*, 46(2), 236-247.
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., & Häussler, T. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2-3), 93-118.
- Owen, M. (2004). An anatomy of games: A discussion paper. Bristol, UK, Future-Lab.
- Papadimitriou, C. H., Prabhakar, R. Tamaki, H., & Vempala, S. (2000). Latent Semantic Indexing: A Probabilistic Analysis, *Journal of Computer and System Sciences*, 61(2), 217-235,
- Phan, M. H., Keebler, J. R., & Chaparro, B. S. (2016). The Development and Validation of the Game User Experience Satisfaction Scale (GUESS). *Human Factors*, 58(8), 1217–1247.
- Poole, S. (2011). *Trigger happy: Videogames and the entertainment revolution*. New York: Arcade Publishing, Inc.
- Strååt, B., & Verhagen, H. (2017). Using user created game reviews for sentiment analysis: A method for researching user attitudes. *Paper presented at the Proceedings of Games-Human Interaction Italy '17 Workshop*.
- Sievert, C., Shirley, K. (2014). LDAvis: a method for visualizing and interpreting topics. *Interactive Language Learning, Visualization, and Interfaces*, 63-70

- Sherrick, B., & Schmierbach, M. (2016). The Effects of Evaluative Reviews on Market Success in the Video Game Industry. *The Computer Games Journal*, 5(3-4), 185-194.
- Tirunillai, S., & Tellis, G. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of Marketing Research*, 51(4), 463-479
- Tong, Z., & Zhang, H. (2016). A Text Mining Research Based on LDA Topic Modelling. *Computer Science & Information Technology*. 6, 201-210.
- Washburn, M., Sathiyarayanan, P., Nagappan M., Zimmermann, T., & Bird, C. (2016). What went right and what went wrong: An analysis of 155 postmortems from game development. *Proceedings of the 38th International Conference on Software Engineering (ICSE)*, IEEE/ACM, 280-289
- 吳尚緯 (2013)。以文字探勘對行動遊戲進行軟體品質特徵分析。國立屏東科技大學工業管理系所碩士論文。屏東縣。取自：
<https://hdl.handle.net/11296/p2amee>
- 簡育銓 (2021)。Newtalk新聞。大發利市！Steam 2020年再創各項紀錄新高。取自：
<https://newtalk.tw/news/view/2021-01-14/523235>

附錄

附錄一 套件導入及文字預處理程式碼

```
# 導入套件
import pandas as pd
import re
from pprint import pprint
from nltk.corpus import stopwords
stop_words = stopwords.words('english')
import gensim
import gensim.corpora as corpora
from gensim.utils import simple_preprocess
from gensim.models import CoherenceModel
from gensim.models.wrappers import LdaMallet # Build LDA model
import os
os.environ.update({'MALLET_HOME':r'C:/Users/chwan/mallet-2.0.8/'})
mallet_path = 'C:\\Users\\chwan\\mallet-2.0.8\\bin\\mallet'
import spacy
# 導入評論文檔/初步處理
df=pd.read_csv("C:\\Users\\chwan\\PycharmProjects\\Essay\\data\\Cyberpunk 2077_reviews.csv")
review = df.review
data = review.values.tolist()
# Remove Emails
data = [re.sub('\S*\S*\s?', '', sent) for sent in data]
# Remove new line characters
data = [re.sub('\s+', ' ', sent) for sent in data]
# Remove distracting single quotes
data = [re.sub("'", "", sent) for sent in data]

def sent_to_words(sentences):
    for sentence in sentences:
        yield(gensim.utils.simple_preprocess(str(sentence), deacc=True))
data_words = list(sent_to_words(data))
# 定義停止詞
def remove_stopwords(texts):
    return [[word for word in simple_preprocess(str(doc)) if word not in stop_words] for doc in texts]

def lemmatization(texts, allowed_postags=['NOUN', 'ADJ', 'VERB', 'ADV']):
    texts_out = []
    for sent in texts:
        doc = nlp(" ".join(sent))
        texts_out.append([token.lemma_ for token in doc if token.pos_ in allowed_postags])
    return texts_out
data_words_nostops = remove_stopwords(data_words)
# 詞型還原
nlp = spacy.load('en_core_web_sm', disable=['parser', 'ner'])
# Do Lemmatization keeping only noun, adj, vb, adv
data_lemmatized = lemmatization(data_words_nostops, allowed_postags=['NOUN', 'ADJ', 'VERB', 'ADV'])
# 建立詞典(Dictionary)與語料庫(Corpus)
id2word = corpora.Dictionary(data_lemmatized)

# 建立詞典(Dictionary)與語料庫(Corpus)
id2word = corpora.Dictionary(data_lemmatized)
# Create Corpus
texts = data_lemmatized
# Term Document Frequency
corpus = [id2word.doc2bow(text) for text in texts]
```

附錄二 一致性分數計算程式碼

```
model_list, coherence_values = compute_coherence_values(dictionary=id2word,
                                                       corpus=corpus,
                                                       texts=data_lemmatized,
                                                       start=2,
                                                       limit=40,
                                                       step=6)

import pyLDAvis
import pyLDAvis.gensim_models # don't skip this
import matplotlib.pyplot as plt
# 輸出最適主題數量圖
limit=40; start=2; step=6;
x = range(start, limit, step)
plt.plot(x, coherence_values)
plt.xlabel("Num Topics")
plt.ylabel("Coherence score")
plt.legend(("coherence_values"), loc='best')
plt.show()
for m, cv in zip(x, coherence_values):
    print("Num Topics =", m, " has Coherence Value of", round(cv, 4))
```

附錄三 LDA分析程式碼

```
# 建構LDA主題模型
lda_model = gensim.models.ldamodel.LdaModel(corpus=corpus,
                                           id2word=id2word,
                                           num_topics=14,
                                           random_state=100,
                                           update_every=1,
                                           chunksize=100,
                                           passes=10,
                                           alpha='auto',
                                           per_word_topics=True)

# 查看LDA模型的主題
pprint(lda_model.print_topics())
doc_lda = lda_model[corpus]
# 計算一致性分數
coherence_model_lda = CoherenceModel(model=lda_model,
                                     texts=data_lemmatized,
                                     dictionary=id2word,
                                     coherence='c_v')
coherence_lda = coherence_model_lda.get_coherence()
print('\nCoherence Score: ', coherence_lda)

# 使用MALLET套件建構LDA模型
ldamallet = gensim.models.wrappers.LdaMallet(mallet_path,
                                             corpus=corpus,
                                             num_topics=14,
                                             id2word=id2word)

# 查看LDA MALLET模型的主題
pprint(ldamallet.show_topics(formatted=False))
# 計算一致性分數
coherence_model_ldamallet = CoherenceModel(model=ldamallet,
                                           texts=data_lemmatized,
                                           dictionary=id2word,
                                           coherence='c_v')
coherence_ldamallet = coherence_model_ldamallet.get_coherence()
print('\nCoherence Score: ', coherence_ldamallet)
```

附錄四 電馭叛客2077主題列表

```
[(0,
  '0.026*"weapon" + 0.024*"level" + 0.020*"gun" + 0.020*"system" + '
  '0.020*"combat" + 0.019*"enemy" + 0.017*"craft" + 0.014*"find" + '
  '0.014*"skill" + 0.013*"hack"'),
(1,
  '0.123*"world" + 0.052*"open" + 0.051*"game" + 0.047*"rpg" + '
  '0.041*"cyberpunk" + 0.023*"expect" + 0.023*"gta" + 0.018*"fallout" + '
  '0.016*"witcher" + 0.012*"action"'),
(2,
  '0.094*"run" + 0.066*"pc" + 0.035*"graphic" + 0.033*"high" + 0.032*"fps" + '
  '0.029*"setting" + 0.024*"issue" + 0.024*"performance" + 0.022*"console" + '
  '0.020*"play"'),
(3,
  '0.049*"release" + 0.037*"year" + 0.021*"promise" + 0.017*"red" + '
  '0.016*"developer" + 0.015*"early" + 0.014*"work" + 0.013*"hype" + '
  '0.012*"devs" + 0.012*"refund"'),
(4,
  '0.069*"play" + 0.066*"people" + 0.044*"love" + 0.044*"enjoy" + 0.032*"game" + '
  '+ 0.031*"review" + 0.026*"thing" + 0.021*"experience" + 0.019*"make" + '
  '0.019*"honestly"'),
(5,
  '0.045*"hour" + 0.036*"work" + 0.035*"crash" + 0.032*"save" + 0.026*"play" + '
  '0.022*"start" + 0.021*"back" + 0.020*"time" + 0.018*"issue" + 0.016*"hard"'),
(6,
  '0.058*"feel" + 0.023*"ai" + 0.018*"lack" + 0.018*"system" + 0.018*"feature" + '
  '+ 0.016*"miss" + 0.016*"bad" + 0.014*"mechanic" + 0.013*"gameplay" + '
  '0.013*"content"'),
(7,
  '0.110*"story" + 0.088*"quest" + 0.088*"side" + 0.066*"mission" + '
  '0.065*"main" + 0.025*"feel" + 0.025*"complete" + 0.022*"line" + 0.020*"end" + '
  '+ 0.020*"finish"'),
(8,
  '0.054*"city" + 0.042*"cyberpunk" + 0.033*"night" + 0.019*"feel" + '
  '0.009*"find" + 0.009*"life" + 0.009*"real" + 0.008*"detail" + 0.007*"live" + '
  '+ 0.007*"moment"'),
(9,
  '0.269*"bug" + 0.054*"glitch" + 0.054*"break" + 0.038*"love" + '
  '0.038*"glitche" + 0.035*"amazing" + 0.035*"recommend" + 0.035*"enjoy" + '
  '0.030*"story" + 0.029*"awesome"'),
(10,
  '0.041*"car" + 0.030*"drive" + 0.027*"npc" + 0.016*"vehicle" + '
  '0.015*"police" + 0.015*"shoot" + 0.012*"walk" + 0.010*"spawn" + '
  '0.009*"random" + 0.009*"kill"'),
(11,
  '0.132*"story" + 0.050*"character" + 0.047*"experience" + 0.045*"gameplay" + '
  '0.039*"amazing" + 0.023*"beautiful" + 0.021*"visual" + 0.021*"good" + '
  '0.019*"fantastic" + 0.019*"graphic"'),
(12,
  '0.078*"fix" + 0.076*"bug" + 0.075*"wait" + 0.058*"buy" + 0.057*"patch" + '
  '0.054*"worth" + 0.039*"buggy" + 0.035*"update" + 0.032*"recommend" + '
  '0.029*"issue"'),
(13,
  '0.067*"character" + 0.032*"choice" + 0.028*"end" + 0.023*"feel" + '
  '0.023*"option" + 0.019*"change" + 0.019*"make" + 0.017*"dialogue" + '
  '0.016*"matter" + 0.013*"choose"')]
```

附錄五 巫師3：狂獵主題列表

```
[(0,
  '0.028*"level" + 0.020*"fight" + 0.018*"system" + 0.017*"enemy" + '
  '0.014*"combat" + 0.014*"difficulty" + 0.012*"horse" + 0.012*"sword" + '
  '0.011*"high" + 0.011*"potion)'),
(1,
  '0.027*"life" + 0.021*"back" + 0.016*"stop" + 0.014*"day" + 0.013*"start" + '
  '0.012*"people" + 0.011*"man" + 0.011*"hate" + 0.011*"leave" + 0.010*"real)'),
(2,
  '0.036*"geralt" + 0.034*"monster" + 0.022*"hunt" + 0.018*"wild" + '
  '0.018*"find" + 0.016*"witcher" + 0.009*"part" + 0.008*"place" + '
  '0.008*"people" + 0.007*"kill)'),
(3,
  '0.157*"quest" + 0.085*"story" + 0.079*"side" + 0.062*"main" + 0.039*"gwent" '
  '+ 0.026*"interesting" + 0.022*"explore" + 0.022*"find" + 0.021*"mission" + '
  '0.020*"card)'),
(4,
  '0.122*"love" + 0.048*"story" + 0.047*"enjoy" + 0.028*"graphic" + '
  '0.027*"skyrim" + 0.025*"mod" + 0.022*"run" + 0.018*"hard" + 0.017*"bug" + '
  '0.016*"word)'),
(5,
  '0.057*"year" + 0.053*"masterpiece" + 0.032*"experience" + 0.029*"red" + '
  '0.026*"game" + 0.022*"projekt" + 0.021*"cyberpunk" + 0.020*"release" + '
  '0.019*"developer" + 0.017*"gaming)'),
(6,
  '0.050*"witcher" + 0.040*"choice" + 0.032*"character" + 0.029*"make" + '
  '0.028*"book" + 0.025*"read" + 0.019*"series" + 0.016*"decision" + '
  '0.016*"feel" + 0.014*"previous)'),
(7,
  '0.241*"game" + 0.177*"play" + 0.105*"good" + 0.088*"make" + 0.035*"single" '
  '+ 0.034*"great" + 0.033*"perfect" + 0.032*"favorite" + 0.031*"player" + '
  '0.024*"top)'),
(8,
  '0.070*"combat" + 0.059*"feel" + 0.026*"bit" + 0.018*"voice" + '
  '0.017*"control" + 0.015*"act" + 0.013*"boring" + 0.013*"write" + '
  '0.013*"dialogue" + 0.013*"review)'),
(9,
  '0.225*"story" + 0.172*"amazing" + 0.087*"recommend" + 0.081*"gameplay" + '
  '0.077*"graphic" + 0.043*"awesome" + 0.039*"fantastic" + 0.034*"highly" + '
  '0.024*"line" + 0.019*"absolutely)'),
(10,
  '0.051*"story" + 0.048*"finish" + 0.043*"hour" + 0.035*"end" + 0.033*"start" '
  '+ 0.032*"dlcs" + 0.031*"dlc" + 0.025*"blood" + 0.024*"expansion" + '
  '0.023*"wine)'),
(11,
  '0.131*"buy" + 0.087*"hour" + 0.079*"worth" + 0.046*"play" + 0.034*"pc" + '
  '0.033*"sale" + 0.029*"full" + 0.028*"money" + 0.026*"price" + '
  '0.025*"spend)'),
(12,
  '0.177*"world" + 0.070*"open" + 0.064*"character" + 0.027*"beautiful" + '
  '0.020*"detail" + 0.019*"immersive" + 0.016*"fantasy" + 0.016*"engage" + '
  '0.016*"combat" + 0.014*"soundtrack)'),
(13,
  '0.276*"good" + 0.156*"rpg" + 0.155*"play" + 0.030*"rpgs" + 0.029*"witcher" '
  '+ 0.029*"decade" + 0.024*"hand" + 0.022*"epic" + 0.019*"coin" + '
  '0.016*"toss")]
```

附錄六 隻狼：暗影雙死主題列表

```
[(0,
  '0.035*"run" + 0.032*"work" + 0.026*"control" + 0.025*"pc" + '
  '0.025*"controller" + 0.024*"issue" + 0.015*"fix" + 0.014*"fps" + '
  '0.014*"steam" + 0.013*"recommend"'),
(1,
  '0.148*"feel" + 0.049*"make" + 0.045*"sword" + 0.026*"bit" + 0.024*"ninja" + '
  '0.023*"thing" + 0.020*"find" + 0.012*"part" + 0.011*"tenchu" + '
  '0.011*"kind"'),
(2,
  '0.206*"die" + 0.092*"time" + 0.058*"hard" + 0.037*"make" + 0.033*"shadow" + '
  '0.020*"challenge" + 0.019*"life" + 0.016*"perfect" + 0.015*"throw" + '
  '0.015*"real"'),
(3,
  '0.041*"people" + 0.038*"review" + 0.030*"start" + 0.028*"find" + '
  '0.026*"back" + 0.019*"put" + 0.015*"give" + 0.014*"understand" + '
  '0.013*"end" + 0.013*"ill"'),
(4,
  '0.202*"game" + 0.071*"fromsoft" + 0.065*"software" + 0.044*"soulsborne" + '
  '0.036*"title" + 0.028*"previous" + 0.026*"fan" + 0.025*"experience" + '
  '0.022*"bloodborne" + 0.019*"favorite"'),
(5,
  '0.164*"hard" + 0.145*"love" + 0.041*"awesome" + 0.041*"hate" + '
  '0.036*"defeat" + 0.032*"cool" + 0.028*"super" + 0.022*"hesitation" + '
  '0.020*"man" + 0.016*"fuck"'),
(6,
  '0.043*"skill" + 0.030*"tool" + 0.029*"weapon" + 0.026*"item" + '
  '0.024*"level" + 0.024*"prosthetic" + 0.018*"find" + 0.018*"character" + '
  '0.017*"lack" + 0.016*"upgrade"'),
(7,
  '0.114*"story" + 0.065*"gameplay" + 0.049*"beautiful" + 0.045*"design" + '
  '0.031*"world" + 0.030*"level" + 0.027*"graphic" + 0.024*"music" + '
  '0.023*"challenge" + 0.022*"interesting"'),
(8,
  '0.114*"amazing" + 0.098*"good" + 0.084*"recommend" + 0.081*"challenge" + '
  '0.062*"fromsoftware" + 0.038*"highly" + 0.032*"masterpiece" + '
  '0.029*"absolutely" + 0.027*"ring" + 0.026*"rewarding"'),
(9,
  '0.089*"enemy" + 0.087*"attack" + 0.034*"deflect" + 0.031*"posture" + '
  '0.026*"block" + 0.022*"damage" + 0.022*"health" + 0.020*"dodge" + '
  '0.020*"hit" + 0.018*"counter"'),
(10,
  '0.131*"play" + 0.087*"buy" + 0.066*"worth" + 0.047*"game" + 0.044*"good" + '
  '0.041*"year" + 0.028*"enjoy" + 0.017*"price" + 0.017*"full" + '
  '0.015*"recommend"'),
(11,
  '0.466*"soul" + 0.255*"dark" + 0.037*"bloodborne" + 0.026*"series" + '
  '0.017*"fan" + 0.015*"expect" + 0.015*"similar" + 0.012*"demon" + '
  '0.011*"compare" + 0.008*"sekiro"'),
(12,
  '0.069*"easy" + 0.065*"game" + 0.060*"play" + 0.053*"hard" + '
  '0.048*"difficulty" + 0.040*"make" + 0.038*"difficult" + 0.028*"enjoy" + '
  '0.022*"video" + 0.018*"people"'),
(13,
  '0.103*"beat" + 0.090*"hour" + 0.064*"feel" + 0.049*"boss" + 0.037*"finally" + '
  '+ 0.034*"finish" + 0.029*"play" + 0.023*"playthrough" + 0.022*"spend" + '
  '0.021*"complete"'),
(14,
  '0.317*"boss" + 0.150*"fight" + 0.032*"mini" + 0.020*"kill" + 0.018*"final" + '
  '+ 0.017*"feel" + 0.016*"end" + 0.014*"area" + 0.013*"make" + '
  '0.012*"cheese"'),
```

```

(15,
'0.115*"learn" + 0.056*"mechanic" + 0.046*"difficult" + 0.036*"master" + '
'0.027*"extremely" + 0.027*"rewarding" + 0.026*"experience" + 0.023*"curve" '
'+ 0.020*"satisfy" + 0.017*"patience"'),
(16,
'0.269*"combat" + 0.100*"system" + 0.046*"fast" + 0.037*"action" + '
'0.027*"pace" + 0.026*"style" + 0.023*"satisfy" + 0.016*"suck" + 0.014*"rpg" '
'+ 0.014*"fluid"'),
(17,
'0.054*"enemy" + 0.030*"kill" + 0.022*"camera" + 0.021*"hit" + 0.015*"run" + '
'0.014*"bad" + 0.014*"move" + 0.013*"stealth" + 0.012*"back" + '
'0.011*"animation"'),
(18,
'0.046*"player" + 0.032*"death" + 0.019*"lose" + 0.018*"point" + '
'0.017*"mechanic" + 0.016*"design" + 0.014*"single" + 0.013*"punish" + '
'0.012*"encounter" + 0.011*"simply"'),
(19,
'0.026*"world" + 0.025*"set" + 0.025*"character" + 0.024*"japanese" + '
'0.024*"story" + 0.015*"player" + 0.010*"fromsoftware" + 0.010*"unique" + '
'0.009*"sekiro" + 0.008*"voice"'])

```

